DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

# THE SCOPE OF THE HYPOTHESIS OF BAYESIAN EQUILIBRIUM

John O. Ledyard


SOCIAL SCIENCE WORKING PAPER 532

June 1984
Revised December 1985

ABSTRACT

What behavior can be explained as the Bayes equilibrium of some game? The main finding is almost anything. Given any Bayesian (coordination) game with positive priors and given any vector of nondominated strategies, there is an increasing transformation of each utility function such that the given vector of strategies is a Bayes (Nash) equilibrium of the transformed game. Any nondominated behavior can be rationalized as Bayes equilibrium behavior. Some comments on the implications of these results for game theory are included.

# THE SCOPE OF THE HYPOTHESIS OF BAYESIAN EQUILIBRIUM

John O. Ledyard*

## I. INTRODUCTION

There has recently been much success in explaining a variety of diverse economic behaviors as outcomes associated with the Bayesian equilibrium of a game. One of the more elegant examples can be found in the use of the revelation principle and the hypothesis that a coordination mechanism be incentive compatible, and efficient, to derive restrictions on the form of optimal auctions. The power of this approach can be seen in Matthews [12], Milgrom and Weber [14], Myerson [17], Myerson and Satterthwaite [18], Wilson [21], Wilson [22] and Gresik and Satterthwaite [4]. In this research, a standard set of simplifying assumptions appears. These frequently include risk-neutral agents with quasi-linear preferences and independently distributed private values. An unanswered question is whether it is the assumption of Bayes equilibrium behavior or the assumptions of specific utility functions and beliefs which drive the results. If the former, then the assumptions are merely simplifying and the conclusions of research in this area can be widely applied; if the latter, then the assumptions are substantive and care must be taken not to attribute too much to any particular result.

To sort this out I simply ask, "what aggregate behavior can be rationalized as the Bayesian equilibrium of some game?" Economists will recognize a close similarity to the question, "what can be an aggregate excess demand function?" (Sonnenschein [20]). I am interested, therefore, in the content of Bayesian equilibrium and incentive compatibility as positive theories. I strongly believe that they are sensible normative principles for guiding the behavior of individuals in strategic situations with incomplete information; but unless the hypothesis of Bayesian equilibrium provides implications independently of the specific functional form of utilities and beliefs, it may be of little use as a positive model to explain actual observations. In particular, it would then be a mistake to identify certain observed behavior or institutions as "not sensible."

The basic components of models which utilize Bayesian analysis are admirably detailed in Myerson [17] to which I refer the novice reader. These components are agents, their preferences over outcomes, their beliefs, and a mechanism for converting messages of the agents into outcomes. Consistent with Bayesian analysis, it is also assumed that beliefs can be represented as probability measures and behavior can be derived from maximization of expected utility functions which are consistent with both preferences and beliefs. The content of the Bayesian equilibrium hypothesis obviously depends on whether or not these components are restricted a priori to some admissible subset. To make a systematic inquiry into the explanatory power of that hypothesis, we will successively tighten the a priori restrictions on

these components and identify what behavior remains consistent with the hypothesis. As we will see, the restrictions must be very severe before any meaningful implications for behavior arise.

I begin, in Section III, by applying a slight variation of the revelation principle, due to Gibbard [3], which significantly simplifies the analysis. Referring to the combination of preferences, beliefs, and utilities as the environment (Hurwicz [7]), we observe that, subject to some informational restrictions, behavior in the context of a specific mechanism, or game form, is a Bayesian equilibrium in some admissible environment if and only if there is an admissible environment such that the performance function, the composite mapping of the specified strategies and outcome rules, is itself an incentive compatible direct revelation mechanism. This means that if we can identify the class of direct revelation mechanisms for which there is an admissible environment such that that mechanism is incentive compatible, then we will know what observed behavior is consistent with the Bayesian equilibrium hypothesis.

In Section IV.2 we show that without further restrictions on the choice of environment, any performance can be rationalized in a non-trivial way by a Bayesian equilibrium. In Sections IV.3 and IV.4 we show that, even if both prior beliefs and preferences are specified a priori for each vector of types, any reasonable performance can still be rationalized as a Bayesian equilibrium for some utility functions consistent with those preferences. In particular, if and only if truth is a dominated strategy for that performance function



qua mechanism can we not rationalize behavior. (A corollary of this result is that if we can choose utilities, the only behavior we cannot rationalize is the behavior of agents who do not use a dominant strategy when one is available.)

This key result can be stated more starkly. Given any Bayesian game with positive priors and any strategy which is not dominated, there is a monotonically increasing transformation of the utility functions such that the given strategy is a Bayesian equilibrium of the transformed game. Since the transformation may need to depend on the entire vector of types, we consider, in Section IV.5, restrictions on utilities. Some results are obtained but a full characterization remains to be done.

A summary of all the results and some observations are provided in Section V. For now let me simply state what I consider to be the main finding of this research: even if preferences and beliefs of all agents, as well as the game form, are pre-specified, any non-dominated behavior can be rationalized as Bayesian equilibrium behavior. Apparently arbitrary non-dominated behavior should be eliminated a priori as not being sensible, only if one is willing to assume very specific functional forms for utilities and beliefs.

## II. THE FRAMEWORK OF ANALYSIS

The context of my analysis is the Bayesian collective-choice problem. Although I am convinced that most of my results can be extended to more general Bayesian incentive problems, I have chosen

this context for ease of comparison to other papers. A Bayesian collective-choice problem is an incomplete information game in which outcomes are jointly feasible for all players together. A mechanism is used to provide the collective choice, given information provided by the agents. For a clear presentation of this model and its philosophical foundations see Myerson [17].

More formally, there are n agents numbered $i = 1,\ldots,n$. We let $T^i$ denote the set of possible types of agent i, and let $T = T^1 \times \ldots \times T^n$. D is the set of possible outcomes or group choices. Each agent has a von Neumann-Morgenstern utility function, $u^i(d,t)$, which denotes the payoff to i if d is the group choice and if $t = (t_1,\ldots,t_n)$ is the vector of agents' types. Each agent also has a probability function $p^i(t_{-i},t_i)$, where $t_{-i} \in T_{-i} \equiv T^1 \times \ldots \times T^{i-1} \times T^{i+1} \times \ldots \times T^n$, which denotes the subjective probability that player i would assign to the event $t_{-i}$ if i's actual type were $t_i$.

The tuple $\alpha = [T,D,u^1,p^1,\ldots,u^n,p^n]$ is called a Bayesian collective-choice problem by Myerson. We assume, as is standard, that the structure of $\alpha$ is common knowledge to all players and that each i knows his own type. I will call the 2-tuple of functions, $\langle u^i,p^i\rangle = e^i$, the characteristic of i and the vector $e = \langle e^1,\ldots,e^n\rangle$, the environment. This language is consistent with that pioneered by Hurwicz [7] and in common use in the literature.

To model the method by which outcomes are selected, as a function of players' types, I use the concept of a mechanism (Hurwicz [7]), sometimes called a game-form (Gibbard [3]). A mechanism is a



pair, $\langle M,g\rangle$ where $M = M^1 \times \ldots \times M^n$. The set $M^i$ is the possible messages agent i can use and M is called the language. The outcome rule $g(m^1,\ldots,m^n)$ maps M into probability measures on D. We let $\Delta(D)$ be the set of all such measures and assume the structure of $\langle M,g\rangle$ is common knowledge. This model of a mechanism is perfectly general and can cover sealed-bid auctions, oral auctions in which players oral responses can depend on others' responses, bargaining models, political processes, etc. Messages can be simple numbers, a monetary bid, or complex conditional responses of the form, if he says a and she says b and then he says c, I will say d. Thus, even though one may not know how to explicitly describe the mechanism of a particular complicated institution such as the experimental Double Oral Auction or the used car market, it is possible to consider them, conceptually, as mechanisms.

Given a mechanism $\langle M,g\rangle$, agents choose messages $m^i$ as a function of their types, and their common knowledge. We call a mapping $\beta^i : T^i \to M^i$ a strategy for i and sometimes use $\beta^i(t^i,g,\alpha)$ to denote its (possible) dependency on the common knowledge of $(g,\alpha)$. Vectors of strategies of particular interest as the fundamental solution concept in this model are the Bayes equilibria for the mechanism $\langle M,g\rangle$ in the collective choice game $\alpha = \langle T,D,e\rangle$. We let $\mu(d,m)$ be the probability assigned to d by the measure $g(m) \in \Delta(D)$. Then

$$\int u^i(d,t)p^i(t_{-i},t_i)\mu(d,\beta^1(t^1),\ldots,\beta^n(t^n))dd\ dt_{-i}$$

represents the expected utility payoff to i if i's type is $t^i$ and if $\beta = \langle \beta^1, \ldots, \beta^n \rangle$ is the vector of strategies used by each player, where $\int$ is the appropriate Riemann-Stieljes integral. Formally, $\beta^* = \langle \beta^{*1}, \ldots, \beta^{*n} \rangle$ is a Bayesian equilibrium of g in $\alpha$, if and only if, for each player i, $\beta^{*i}$ is a function from $T^i$ to $M^i$ such that for each $t^i \in T^i$

$$\int u^i(d,t) p^i(t_{-i}, t_i) \mu(d, \beta^*(t)) \, dd \, dt_{-i} = \max_{m^i \in M^i} \int u^i(d,t) p^i(t_{-i}, t_i) \mu(d, \beta^*(t)/m_i) \, dd \, dt_i$$

where

$$\langle \beta^*(t)/m^i \rangle = \langle \beta^{*1}(t^1), \ldots, \beta^{*i-1}(t^{i-1}), m^i, \beta^{*i+1}(t^{i+1}), \ldots, \beta^{*n}(t^n) \rangle.$$

I will sometimes use $\beta^*(t;\alpha,g)$ to represent the dependency of the Bayes-equilibrium on the common knowledge of $\alpha$ and $\langle M,g \rangle$.

It is convenient to have a way to summarize the net result of the combination of $\alpha$,g and $\beta^*$. This is typically done with a performance function which describes the choice, d, made for each vector of types, t. Formally, we call the function $\Pi : T \to \Delta(D)$, where $\Pi(t) = g(\beta(t))$, the performance of g in $\alpha$ for the strategy $\beta$. Of particular interest is the performance of g in $\alpha$ for the Bayes-equilibrium strategy.[1] We will denote that as $\Pi^*(t) = g(\beta^*(t;g,\alpha))$.

## III. QUESTIONS AND PRELIMINARIES

Two main types of questions have been addressed in this framework. The first type, generally called Optimal Mechanism Design,



asks: Given $\alpha$ if $\beta^*$ is Bayes equilibrium what performance is possible by varying the mechanism? Also, what mechanism gives the best performance? This type of question was posed but unanswered in Hurwicz [8]. Later papers by Harris and Raviv [5], Harris and Townsend [6], Matthews [12], Myerson [16], and Wilson [22], have provided some answers. Wilson's paper provides a good summary. The research rests on the revelation principle (first formalized by Gibbard [3]). Since this principle is of importance to our later results let us take a brief glimpse.

We call any mechanism $\langle M,g \rangle$ a direct-revelation mechanism if $M^i = T^i$ for all agents i. That is, each agent announces a, possibly false, type which is used by $g(t^1, \ldots, t^n)$ to pick $\mu \in \Delta(D)$. Of particular interest are direct-revelation mechanisms for which truth is a reasonable strategy. We call a mechanism $\langle M,g \rangle$ an incentive-compatible direct-revelation mechanism for $\alpha$ (an icdr) if and only if $M^i = T^i$ for all agents and $\beta^{*i}(t^i) = t^i$, for all $t^i \in T^i$ and all i, is a Bayes equilibrium for g in $\alpha$.

[Note: This does not require that $\beta^*$ be the unique Bayes equilibrium. See Postlewaite-Schmeidler [19] for some of the problems of non-uniqueness.]

Theorem 1: (Revelation Principle, Gibbard [3]) Given $\langle M, g \rangle$, $\alpha$, and $\beta^i : T^i \to M^i$, for each i, such that $\beta^i(T^i) = M^i$.

A. $\beta$ is a Bayes equilibrium for $\langle M,g \rangle$ in $\alpha$

if and only if

B. $\langle T,\sigma\rangle$, where $\sigma(t) = g(\beta(t))\ \forall t$, is an icdr in $\alpha$.

Remark: The condition that $\beta^i(T^i) = M^i$ is needed to show that B $\Rightarrow$ A. The statement that A $\Rightarrow$ B is valid even if $\beta^i(T^i) \subset M^i$ .

Thus one answer to the question, what performance is possible, is that any function from T to $\Delta(D)$ which is itself an icdr is a possible performance function for some mechanism and only those functions are possible. The question, what mechanism gives the best performance, involves choosing among all icdr's. The standard method is to define a concept of efficiency and then to select an efficient icdr. Since we will not need that machinery for this paper, we refer the interested reader to Myerson [17].

The second type of common question has been concerned with the implementation of desired performance. It has been asked: given $\alpha$ and $\Pi$, does there exist a mechanism $\langle M,g\rangle$ such that, if $\beta^*$ is a Bayes equilibrium for g in $\alpha$ then $g(\beta^*(t)) = \Pi(t)$? This question has been addressed by Laffont and Maskin [9], Postlewaite and Schmeidler [19], and Ledyard [10], among others. Using the Revelation Principle, the answer is yes only if $\Pi$ is itself an icdr mechanism.[2] The converse needs some additional assumptions. See Theorem 2 below.

One can argue that much of the above is an essentially normative approach to behavior in strategic situations with differential information. While the answers to the first question, what can be the performance of some g in $\alpha$, give some direction to the



positive construction of explanatory models, they do not go far enough. We need to know the answer to the following question: given T, D, $\langle M,g\rangle$ and $\Pi$ does there exist $\alpha$ such that if $\beta^*$ is a Bayes-equilibrium for g in $\alpha$ then $g(\beta^*(\ )) = \Pi(\ )$. In effect, given the institution $\langle M,g\rangle$, can we "rationalize" the performance $\Pi$ as the result of Bayesian behavior?[3]

Since I have already indicated in the introduction why one should be interested in this question, let me turn to an initial answer which is closely related to the revelation principle.

Theorem 2: Given $\alpha$, $\Pi$, and $\langle M,g\rangle$.

A. $\beta^i : T^i \to M^i$, for each i, such that:

A.1: $\beta^i(T^i) = M^i$, for each i,

A.2: $\beta$ is a Bayes-equilibrium for $\langle M,g\rangle$ in $\alpha$, and

A.3: $\Pi(t) = g[\beta(t)] \quad \forall t \in T$,

if and only if

B. The following are true:[4]

B.1: $[m|g(m) = \Pi(t)] \neq \emptyset, \ \forall t \in T$,

B.2: There is a selection

$$[\eta^1(t^1),\ldots,\eta^n(t^n)] \in [m|g(m) = \Pi(t)] \quad \forall t$$

with the property that $\eta^i(T^i) = M^i \quad \forall i$, and

B.3: $\langle T,\Pi\rangle$ is an icdr mechanism in $\alpha$.

Proof: (A $\Rightarrow$ B) A.3 implies B.1, since then

$$\beta(t) \in \{m \mid g(m) = \Pi(t)\}, \forall t \in T.$$

To establish B.2, let $\eta^i(t^i) = \beta^i(t^i)$ for all $t^i \in T^i$. A.1 and A.3 then imply B.2.

B.3 follows, using theorem 1, from A.2.

(B $\Rightarrow$ A) Let $\beta^i(t^i) = \eta^i(t^i)$ $\forall t^i \in T^i$ and $\forall i$. Then $\eta^i(T^i) = M^i$ implies A.1. Also B.2 implies that $g(\eta(t)) = \Pi(t)$ for each t which implies A.3. Finally using theorem 1 and the fact that $\beta^i(T^i) = M^i$, B.3 implies A.2.

(QED)

Remark: As in Theorem 1, the range conditions that $\eta^i(T^i) = M^i$ and A.1 are only needed to show that B $\Rightarrow$ A. The opposite implication that A $\Rightarrow$ B can be established without these.

This Theorem is mainly the revelation principle with condition B.2 added to ensure that the performance correspondence can be "decomposed" into individual behavioral components consistent with $\langle M,g\rangle$. It is similar, but not quite identical, to the requirement that $\{m \mid g(m) = \Pi(t)\}$ be a coordinate correspondence from T to M in the sense of Mount and Reiter [15].

A corollary to Theorems 1 and 2 provides the main principle of analysis in modeling and rationalizing observed behavior.

Corollary 1.1: Given D,T, $\langle M,g\rangle$, and $\delta^i : T^i \to M^i$, for all i, such that $\delta^i(T^i) = M^i$. $\exists$ $\alpha$ such that $\delta$ is a Bayes equilibrium for g in $\alpha$ if and only if $\exists$ $\alpha$ such that $\langle T,g(\delta(\cdot))\rangle$ is an icdr mechanism in $\alpha$.



If we observe the behavior $\delta$ in the institutional setting described by $\langle M,g\rangle$, we will be able to say agents acted as if they were playing a Bayesian game if and only if there is some environment $\alpha$ such that the performance mapping $\Pi(t) \equiv g[\delta(t)]$ is itself an icdr mechanism in that environment.

The conclusion of this section is obvious. Leaving aside issues of complexity, computability, misperceptions, and the size of the message space, most remaining modeling questions will be answered once we know for which pairs, $\langle\alpha,\Pi\rangle$, of environments and performance maps, $\langle T,\Pi\rangle$ is an icdr mechanism in $\alpha$. I turn to that now.

## IV. INCENTIVE COMPATIBLE DIRECT REVELATION

When can we find an environment, $\alpha$, such that the direct revelation mechanism $\Pi$ is incentive compatible in $\alpha$? Or, when can we rationalize the performance $\Pi$? The answers, provided in this section, depend on the prior restrictions one places on the allowable choices of $\alpha$. As we will see, unless utilities and priors, $\langle u^i,p^i\rangle$, are constrained to lie in a very narrow set, virtually any $\Pi$ is incentive compatible in some allowable $\alpha$.

### IV.1. No a priori restrictions

Our first result is that anything is rationalizable as Bayes Equilibrium behavior. This is trivial but is included for completeness and to provide motivation for the more complicated later propositions.

Theorem 3: Given $\langle D,T\rangle$. There is an e such that every function from T to D is a Direct Revelation Incentive Compatible Mechanism in $\alpha = \langle D,T,e\rangle$.

Proof: For every agent i, and every type $t^i \in T^i$, define $u^i(d,t) = a^i(t^i)$ for all $(d,t_{-i})$ where $a^i(\cdot)$ is any function from $T^i$ to $R^1$. Let $p^i(\cdot,t_i)$ be any measure on $T_{-i}$. Then, since i is indifferent among outcomes, anything, including $\beta^i(t^i) = t^i$, is a Bayes equilibrium for $\langle T,\Pi\rangle$ in this environment, for any mapping $\Pi : T \to \Delta(D)$

(QED)

## IV.2. Any Certainty Mechanism Can Be Rationalized

The previous theorem is not an entirely fair test of the power of the hypothesis of Bayesian equilibrium behavior. There are two obvious ways to make the construction of the required $\alpha$ more difficult: restrict the choice of $u^i$ and $p^i$ or strengthen the concept of incentive compatibility. We will try the latter first.

Definition: A direct revelation mechanism $\Pi : T \to \Delta(D)$ is strictly incentive compatible in $\alpha$ if and only if, for all i, $t^i \in T^i$, $s \in T^i$ such that $\mu(d,t) \neq \mu(d,t/s)$ for some $(d,t_{-i})$,

$$U^i(\alpha,s,t^i) < U^i(\alpha,t^i,t^i) \text{ for } s \neq t^i$$

where

$$U^i(\alpha,s,t^i) = \int u^i(d,t)\mu(d,t/s)dd\ dt_{-i}$$



and $\mu(d,t)$ is the probability assigned to d by the measure $\Pi(t)$.

For strict incentive compatibility, truth must be a unique maximizer of expected utility, unless the mechanism doesn't care. The mechanism can be indifferent, but an agent can not.

Remark: The existence of e such that $\Pi$ is a strictly icdr mechanism in (D,T,e) when D and T are finite, is equivalent to the existence of vectors $Z^i_\ell$ for all i and all $\ell \in T^i$ such that, for all $\ell \in T^i$,

$$\sum_{t-i}\sum_{d} Z^i_\ell(d,t_{-i})[\mu(d,t) - \mu(d,t/s)] > 0$$

when $t = (t/\ell)$ and $\exists(d,t)$ such that $\mu(d,t) \neq \mu(d,t/s)$. This is true because, given $Z^i_\ell(d,t_{-i})$, we can easily find $u^i(d,t)$ and $p^i(t_{-i},\ell)$ such that

$$u^i(d,t/\ell)p^i(t_{-i},\ell) = Z^i_\ell(d,t_{-i}) \quad \forall d,t.$$

Myerson [17] contains a full development of some of the implications of this insight.

We can turn to the theory of linear inequalities for our answers. In this case we use a variation on Farkas' Lemma.

Theorem 4: (Gale [2], Theorem 2.10) If D and T are finite, $\exists Z$ such that

$$\sum_{t_{-i}}\sum_{d} Z(d,t_{-i})[\mu(d,t_{-i},\ell) - \mu(d,t_{-i},s)] > 0$$

for all $s \in T^i$ if and only if there do not exist $x_s \geq 0$, $x \neq 0$ such that

$$\sum_{s \neq \ell} x_s[\mu(d,t_{-i},\ell) - \mu(d,t_{-i},s)] = 0 \qquad (1)$$

for all $(d,t_{-i}) \in D \times T^{-i}$.

Many of our results will be more easily interpreted for mechanisms which only assign probability 0 or 1 to any $d \in D$.

Definition: A direct revelation mechanism $\Pi : T \to D$ is called a c-mechanism (certainty mechanism). Those mechanisms such that $\Pi : T \to \Delta(D)$ which are not c-mechanisms are called p-mechanisms (probabilistic mechanisms).

Remark: Sometimes concentration of the analysis on c-mechanisms can be done without loss of generality. This is because icdr mechanisms form a polyhedron, a closed convex set, given $\alpha$, which can be described by its extreme corners, where $\mu(d,t) = 0$ or 1. However, since we begin with the mechanism and not $\alpha$, we must consider p-mechanisms as well as c-mechanisms.

Corollary 4.1: Given any c-mechanism, $\Pi : T \to D$, with T and D finite $\exists e$ such that $\Pi$ is strictly incentive compatible in $\alpha = (D,T,e)$.

Proof: Pick i and $\ell \in T^i$. We eliminate any s for which $\Pi(t/\ell) = \Pi(t/s)$ for all $t_{-i} \in T^{-i}$. Then, applying Theorem 4, suppose the desired Z vector does not exist. But then $\exists x \geq 0$ such that $\sum_s x_s[\mu(d,t/\ell) - \mu(d,t/s)] = 0$ and $x \neq 0$.

For each s',



$\exists \hat{t}_{-i}$ such that $\Pi(\hat{t}/s') = \hat{d} \neq g(\hat{t}/\ell)$. Thus

$$\sum_s x_s[\mu(\hat{d},t/\ell) - \mu(\hat{d},t/s)] = \sum_s x_s[0 - \mu(\hat{d},t/s)]$$
$$= -x_{s'} - \sum_{s \neq s'} x_s \mu(\hat{d},t/s) = 0.$$

This is true only if $x_{s'} = 0$. Therefore the only solution has $x = 0$ which is a contradiction.

(QED)

Corollary 4.2: Given any p-mechanism $\Pi : T \to \Delta(D)$ with T and D finite such that $\forall i$, $t_i \in T^i$, $s \in T^i$, $\exists (\hat{d},\hat{t}_{-i}) \in D \times T^{-i}$ such that either

$$\mu(\hat{d},\hat{t}) = 1 > \mu(\hat{d},\hat{t}/s) \text{ or } \mu(\hat{d},\hat{t}) = 0 < \mu(\hat{d},\hat{t}/s).$$

$\exists \alpha = (D,T,e)$ such that P is strictly incentive compatible in $\alpha$.

Proof: Essentially the same as Corollary 4.1.

(QED)

Remark: The condition in Corollary 4.2 is a sufficient condition which is not close to being necessary. I have no easily interpretable necessary and sufficient condition.

Remark: It is possible to make precise the statement that almost all direct revelation mechanisms (including p-mechanisms) are rationalizable.

Corollary 4.3: Given D, T, finite, let

$$\mathbb{P} \equiv \{\mu : D \times T \to [0,1] \mid \sum_d \mu(d,t) = 1 \quad \forall t \in T\}$$

be the set of all possible direct revelation mechanisms. $\exists$ an open dense subset $\mathbb{P}^*$ of $\mathbb{P}$ such that if $\mu \in \mathbb{P}^*$ then $\exists e$ such that $\mu$ is strictly incentive compatible in $\alpha = (D,T,e)$.

Proof: Write equation 1 from Theorem 4 as $xA = 0$. A sufficient condition for the existence of the desired e is that the rank of A equal $||T^i - 1|| \equiv \tau$ since then $x = 0$ is the only solution of (1). It is easy to show that if rank (A) $< \tau$ then one can perturb those values of $\mu$ such that $\mu(d,t) \in (0,1)$ and generate A* such that rank (A*) $= \tau$. (Remember if $\mu(d,t) \in [0, 1]$ there is no problem.) One can also do this so that $\sum_d \mu^*(d,t) = 1$ for all t.

(QED)

IV.3. Restrictions on Preferences

One can argue that the fact that any c-mechanism and almost any p-mechanism can be rationalized is not a very reasonable test of the hypothesis of Bayes equilibrium behavior since for most analyses there is a natural order on outcomes induced by the physical description of types and outcomes. That is, the abstractions are not entirely independent of the events being modeled.

For example, in a private values auction, it is standard to let $T^i = W^i \times V^i$ where $v^i \in V^i$ is the monetary value of the object to buyer i. The other component $w^i \in W^i$ is used to represent uncertain characteristics such as risk attitudes, initial wealth, etc. In this case it is assumed that utility is positively increasing in ex post income. That is, $u(d,t) > u(d',t')$ if $v^i y^i - x^i > v^i y'^i - x'^i$ where



$d = \langle d^1,\ldots,d^n\rangle$, $d^i = \langle y^i, x^i\rangle$, $y^i$ is the number of units received, and $x^i$ is the amount to be paid by i. For examples, see Milgrom and Weber [14], and Myerson and Satterthwaite [18].

There are several ways to use the "natural order" to restrict our choice of models. I will concentrate on one. Since the natural order is to be imposed on d for a given t, we will assume we are given, a priori, a function $w^i(d,t)$ which generates that order.

Definition: Given D,T, $w^1(d,t),\ldots,w^n(d,t)$ we say that $e = \langle u^1,p^1,\ldots,u^n, p^n\rangle$ is consistent with $w = \langle w^1,\ldots,w^n\rangle$ if and only if, for all $\hat{d}$, d, and t, $[w^i(d,t) > w^i(\hat{d},t)]$ if and only if $[u^i(d,t) > u^i(\hat{d},t)]$.

Note that only the preference ordering over certainty outcomes, not the preference ordering over lotteries, is preserved, and it is not required that the ordering across vectors of types be preserved. For example, if $w^i(d,t) = w^i(d,t_i)$ (that is w is independent of $t_{-i}$) then consistency requires that $u^i(d,t_{-i},t_i^*)$ and $u^i(d,\hat{t}_{-i},t_i^*)$ order D in the same way, although $u^i$ may still be sensitive to the value of $t_{-i}$. One possibility allowed is that

$$u^i(d,t) = f^i[w^i(d,t_i),t], \text{ where } f^i_1 > 0 \text{ and } f^i_t \neq 0.$$

We can now provide the central result.

Theorem 5 Let D, T be finite. Let $\Pi : T \to \Delta(D)$ be a direct revelation mechanism, where $\mu(d,t)$ is the probability assigned to d by $\Pi(t)$. Let the orders $w = \langle w^1(d,t),\ldots,w^n(d,t)\rangle$ be given. There

exists $e^i = \langle u^i, p^i \rangle$ for each i such that e is consistent with w, $p^i(t) > 0$ $\forall t$, and such that $\Pi$ is an icdr mechanism in $\alpha = (D,T,e)$) if and only if, for every i, there do not exist $\alpha^i_{\ell s} \geq 0$ $\forall \ell$, $s \in T^i$ such that

$$\sum_{s \neq \ell} \alpha^i_{\ell s} \sum_{d' \leq d} \left[ \mu(d', t/\ell) - \mu(d', t/s) \right] \geq 0 \qquad (2)$$

for all $d \in D$ and all $t = (t_{-i}, \ell) \in T$, with strict inequality for one (d,t), where $d' \leq d$ if and only if $w^i(d', t/\ell) \leq w^i(d, t/\ell)$.

Proof: The strategy of the proof is to first show that the existence of the desired environment e is equivalent to the existence of a solution to a set of linear inequalities. Then Farkas' Lemma is applied to get (2).

(i) Incentive compatibility as linear inequalities

Because of the structure of the problem we can treat the proof separately for each $t^i \in T^i$ and for each i. Think of $\mu$ as an $r = L \times K$ dimensional vector where $L = ||T||$ and $K = ||D||$. For any utility, $u^i(d,t)$, and beliefs, $p^i(t)$, let $Z^i(d,t) = u^i(d,t)p^i(t)$. The vector $Z^i \in \mathbb{R}^r$. Now, given $\ell \in T^i$, let $Z^i_\ell$ be the $r - ||T^i|| = r - KL_i$ vector whose components are $Z^i(d,t/\ell)$. Let $Q^i_\ell$ be the $(r - KL_i) \times r$ matrix such that $Z^i_\ell = Q^i_\ell Z^i$. The expected utility received if i uses $\ell$ in the mechanism $\mu$ is

$$U(\mu, \ell | \ell) = Z^i_\ell Q^i_\ell \mu.$$

If, instead, i were to use $s \in T^i$ then i's payoff would be



$$U(\mu, s | \ell) = Z^i_\ell Q^i_s \mu.$$

Thus $\Pi$ is incentive compatible in $\alpha = (D,T,e)$ if and only if

$$Z^i_\ell (Q^i_\ell - Q^i_s)\mu \geq 0 \qquad (3.1)$$

for all i, and all $\ell$, $s \in T^i$.

(ii) e consistent with w as linear inequalities

Given e and w, let $Z^i(d,t) = u^i(d,t)p^i(t)$ as before. For e to be consistent with $w^i$ it must be true that, for all d, t, d',

$$w^i(d,t) > w^i(d',t) \text{ if and only if } u^i(d,t) > u^i(d',t).$$

This is equivalent to, for $p^i(t) > 0$,

$$w^i(d,t)) > w^i(d',t) \text{ if and only if } Z^i(d,t) > Z^i(d',t).$$

There are a variety of ways to represent this constraint on the choice of e. The most straightforward is first to renumber D, given t, so that for $d = 1,\ldots,D'$, $d > d'$ if and only if $w^i(d,t) > w^i(d',t)$ and for $d = D'+1,\ldots,K$ $w^i(d,t) = w^i(d',t)$ for some $d' \leq D'$. Thus, we create a strong order on $d = 1,\ldots,D'$ and number the d indifferent to these with numbers larger than D'. [Since D is finite, this can be done.] We then require that there is $\varepsilon > 0$ such that

$$Z^i(d,t) - Z^i(d-1,t) \geq \varepsilon \quad \forall d, \text{ and } \forall t, \text{ with } 2 \leq d \leq D'.$$

This can be represented in matrix form, where $R^i_t$ is an $r \times (D' - 1)$ matrix for each t and $\varepsilon$ is $1 \times (D' - 1)$, as

$$Z^i R^i_t \geq \varepsilon \text{ for all } t. \tag{3.2}$$

Since $Z^i$ can be multiplied by scalers without affecting incentive compatibility, the restriction that $\varepsilon > 0$ is equivalent to the restriction that $Z^i R^i_t > 0 \quad \forall t$.

For $d = D'+1,\ldots,D$ we require that

$$Z^i(d,t) - Z^i(d',t) \geq 0$$

and

$$Z^i(d',t) - Z^i(d,t) \geq 0$$

where $d'$ is the appropriate number $\leq D'$ such that $w^i(d,t) = w^i(d',t)$. This can also be represented in matrix form, where $I^i_t$ is an $r \times 2(D - D')$ matrix for each t, as

$$Z^i I^i_t \geq 0 \text{ for all } t. \tag{3.3}$$

At this point we know that there is an e consistent with w such that $\Pi$ is incentive compatible in (D, T, e) iff there are vectors $(Z^1,\ldots,Z^n)$ such that equations (3) are true. But this is true iff $\forall i$, $\forall \ell \in T^i$ and $\forall \varepsilon > 0$, $\exists\ Z^i_\ell$ such that

$$Z^i_\ell (Q^i_\ell - Q^i_s)\mu \geq 0 \quad \forall\ s \in T^i \tag{4.1}$$

$$Z^i_\ell R_{t/\ell} \geq \varepsilon \quad \forall t_{-i} \in T^{-i}, \tag{4.2}$$

and

$$Z^i_\ell I^i_{t/\ell} \geq 0 \quad \forall t_{-i} \in T^{-i}, \tag{4.3}$$



where $R_{t/\ell}$ and $I^i_{t/\ell}$ are the appropriate modifications of $R^i_t$ and $I^i_t$.

Applying Farkas' Lemma (see, e.g., Gale [2], Theorem 2.7), $\exists Z^i_\ell$ satisfying (4) if and only if $\nexists \alpha_{\ell s} \geq 0$, for all $s \in T^i$, $\gamma_{d,t/\ell} \geq 0$, for all $1 \leq d \leq D'$, $t_{-i} \in T^{-i}$ and $\delta^1_{d,t} \geq 0$, $\delta^2_{d,t} \geq 0$ for $D' + 1 \leq d \leq K$ and $t_{-i} \in T^i$ such that

$$\sum_{d,t_{-i}} \gamma_{d,t/\ell} \cdot \varepsilon = 1, \tag{5.1}$$

$$\gamma_{1t} = 0 \text{ for all } t_{-i}, \tag{5.2}$$

$$\sum_{s \neq \ell} \alpha_{s\ell}[(Q^i_\ell - Q^i_s)\mu]_{dt/\ell} + \gamma_{d,t/\ell} - \gamma_{d+1,t/\ell} + \sum_{d'=d} (\delta^1_{d',t/\ell} - \delta^2_{d',t/\ell}) = 0 \tag{5.3}$$

for all $d,t_{-i} \in D \times T^{-i}$ with $1 \leq d \leq D'$ where $d' = d$ if and only if $d' > D'$ and $w^i(d',t/\ell) = w^i(d,t/\ell)$, and

$$\sum_{s \neq \ell} \alpha_{s\ell}\,[(Q^i_\ell - Q^i_s)\mu]_{d't/\ell} - (\delta^1_{d',t/\ell} - \delta^2_{d',t/\ell}) = 0 \tag{5.4}$$

for all $d' > D'$, all $t_{-i} \in T^{-i}$.

Solving (5.3) and (5.4) for $\gamma_{d+1,t}$ and letting $\geq$ represent the ordering derived from $w^i$, we get

$$\gamma_{d+1,t/\ell} = \sum_{\hat{d} \leq d} \sum_{s \neq \ell} \alpha_{\ell s}\ [(Q^i_\ell - Q^i_s)\mu]_{\hat{d}t/\ell}$$

for all $d,t_{-i}$, $d \leq D'$. Since $[(Q^i_\ell - Q^i_s)\mu]_{\hat{d}t/\ell} = \mu(\hat{d},t/\ell) - \mu(\hat{d},t/s)$ and since $\gamma_{d,t/\ell} \geq 0$ for all $d,t_{-i}$ with strict inequality for at least

one $d,t_{-i}$, the theorem is proven.

(QED)

Remark 5: If, in the statement of Theorem 5, we replace "icdr" with "strictly icdr" and replace "$\geq 0$" with "$> 0$" in equation (2), the statement remains true.

If we restrict out attention to c-mechanisms, a very natural intuitive necessary and sufficient condition can be detailed. But first we need to define another property of direct revelation mechanisms.

Definition: Given $\Pi : T \to D$ and preferences $w = (w^1,\ldots,w^n)$ where $w^i : D \times T \to R^1$. We say that truth is dominated in $\Pi$ for w if there is an i, a $t^i \in T^i$, and $s \in T^i$, with $t^i \neq s$ such that

$$w^i[\Pi(t/s),\ t_{-i},t_i] \geq w^i[\Pi(t),\ t]$$

for all $t_{-i} \in T^{-i}$, with strict inequality for at least one $t_{-i} \in T^{-i}$. We say that truth is weakly dominated in $\Pi$ for w if there is an i a $t^i \in T^i$, and $s \in T^i$, with $t^i \neq s$ such that

$$w^i[\Pi(t/s),t_{-i},t^i] \geq w^i[\Pi(t),\ t]\text{'for all } t^{-i} \in T^{-i}$$

That is, truth is dominated in $\Pi$ for w if there is some type $t^i$ for whom there is a strategy which guarantees a better outcome than $t^i$. An example of such a mechanism and preferences is given after the next result.



Corollary 5.1: Given the c-mechanism $\Pi : T \to D$ and preferences w, with T and D finite, (a) $\exists a = (D,T,e)$ such that e is consistent with w and $\Pi$ is an icdr mechanism in a if and only if truth is not dominated in $\Pi$ for w. (b) $\exists a = (D,T,e)$ such that e is consistent with w and $\Pi$ is a strictly icdr mechanism in a if and only if truth is not weakly dominated in $\Pi$ for w.

Proof: We prove (a). (b) can be proved in a similar way. (if) If $w^i[\Pi(t),t] = w^i[\Pi(t/s),t]$ for all $t_{-i}$ then (2) holds with equality for all $d,t_{-i}$ and we are done. If not, $\exists t_{-i} \in T^{-i}$ such that $w^i[\Pi(t),t] > w^i[\Pi(t/s),t]$. Let $d = \Pi(t/s)$. Then

$w^i[\Pi(t),t] > w^i(d,t) \geq w^i[\Pi(t/s),t]$. Let

$B = \{s' \in T^i | w^i(\Pi(t/s'),t) \geq w^i(d,t)\}$

Equation (2) of Theorem 5 for $(d,t_{-i})$ becomes $-\sum_{s\in B}\gamma^i(t^i,s) \geq 0$, where $\gamma^i(t^i,s)$ is the appropriate $\gamma$. Therefore, $\gamma^i(t^i,s) = 0$ for all $s \in B$. Continue for all i, all $s \in T^i$ and all $t^i$. It must be true that $\gamma^i_{\ell s} = 0$ $\forall i, \ell, s \in T^i$. Therefore, strict inequality is untrue for any d, $t_{-i}$.

(only if). Suppose truth is dominated in $\Pi$ for w. There is i and $t^i$, $s \in T^i$ such that $w^i(\Pi(t/s),t) \geq w^i[\Pi(t),t]$ for all $t_{-i}$ with strict inequality for one $t_{-i}$. Let $\gamma^k_{\ell s} = 0$ for all k, $\ell$, s except these i, s, $t^i$. Let $\gamma^i_{t^i,s} = 1$. For all k and for i when $\ell \neq t^i$, (2) of Theorem 5 is trivially true. For i, $t^i$, s, let $\forall d,\ t_{-i}$,

$$f^i(d,t_{-i},s,t_i) = \gamma_{t^i s}[\eta^i(d,t,t) - \eta^i(d,t/s,t)].$$

Where

$$\eta^i(d,\ t/s,\ t) = \begin{cases} 0 & \text{if } w^i[\Pi(t/s),t] \geq w^i(d,t) \\ 1 & \text{otherwise} \end{cases}$$

Now if $w^i[\Pi(t/s),t] \leq w^i(d,t)$ then $w^i[\Pi(t),t] \leq w^i(d,t)$. Therefore, $f^i(d,t_{-i},s,t_i) \geq 0$. It is strictly greater for one $t_{-i}$, d since $w^i[\Pi(t),t] < w^i[\Pi(t/s),t]$ for one $t_{-i}$. (Let $d = \Pi(t)$.) Thus $\exists\ \gamma \geq 0$ satisfying (2).

(QED)

We can rationalize the observed behavior if and only if truth is never a dominated strategy for any type. To see what type of restriction this is, consider the following situation which often arises in the early rounds of many experiments. Dominant strategies are available but not used. For example, many times subjects will bid less than their true values in sealed-bid second price auctions. In these, the highest bidder receives the item and pays an amount equal to the second highest bid. It is a dominant strategy to bid your true value. Can we explain this phenomenon? The answer is no.

Corollary 5.2: Given finite D, T, preferences w, and $\langle M,g\rangle$ such that $\delta^i(t^i) = m^i$ is a strictly dominant strategy; that is,

$$w^i[g(m/\delta^i(t^i)),t] \geq w^i[g(m),t]$$



for all $m \in M$, with strict inequality for at least one $m \in M$. Let $\beta^i(t^i) = m^i$ be the observed behavior such that, for some i, $\beta^i(t^i) \neq \delta^i(t^i)$ for some $t^i \in T^i$ and such that $\{\hat{t}^i | \delta^i(\hat{t}^i) = \beta^i(t^i)\} \neq \emptyset$ for all $t^i \in T^i$. Let $\Pi(t) = g(\beta(t))$. $\beta$ cannot be rationalized; that is , $\nexists e$ consistent with w such that $\Pi$ is an icdr mechanism in $\alpha = (d,T,e)$.

Proof: We know that

$$w^i[g(\beta(t)/\delta^i(t^i)),t^i] \geq w^i[g(\beta(t)),t^i]$$

for all $t_{-i}$ with strict inequality for at least one. Let

$\hat{t}^i \in \{s | \beta^i(s) = \delta^i(t^i)\}$. Then

$$w^i[\Pi(t/\hat{t}^i),t^i] \geq w^i[\Pi(t),t^i]$$

for all $t_{-i}$ with strict inequality for at least one $t_{-i}$. Thus truth is dominated in $\Pi$ for w.

(QED)

Individuals who do not follow a dominant strategy, when one is available, can not be rationalized. When there is no dominant strategy, most behavior can be rationalized.

Corollary 5.3: Given finite D, T, preferences w, mechanism $\langle M,g\rangle$, and observed behavior $\beta^i : T^i \to M^i$ for each i such that $\beta^i(T^i) = M^i$. $\exists e$ consistent with w such that $\beta$ is a Bayes equilibrium of $\langle M,g\rangle$ in $\alpha = (D,T,e)$ if and only if, $\forall i$, $\beta^i$ is not a dominated strategy; that

is, for each $t^i \in T^i$ there is no $m^i \in M^i$ such that $w^i[g(\beta(t)),t] \leq w^i[g(\beta(t)/m^i),t]$ for all $t_{-i} \in T^{-i}$ with strict inequality for one $t_{-i}$.

Proof: Let $\Pi(t) = g(\beta(t))$. $\beta$ not dominated if and only if truth is not dominated in $\Pi$.

(QED)

In words, if and only if some type of some agent picks a message which is dominated by some other, can we not rationalize behavior. Any behavior responsive to sure things can be rationalized.

A simple example with which to illustrate the results arises in the context of a one buyer-one seller bilateral bargaining game (Myerson-Satterthwaite [18]). In this game

$$w^B(d,t) = v^B y - x \text{ where } y = \begin{cases} 1 & \text{if buyer gets the item} \\ 0 & \text{otherwise} \end{cases}$$

and x is the monetary transfer to the seller. A direct revelation mechanism is $\Pi(t) = [y(t), x(t)]$. Truth is dominated iff $\exists v^B$ and $Z \in T^B$ such that

$$v^B y(v^B,v^S) - x(v^B,v^S) \geq v^B y(Z,v^S) - x(Z,v^S)$$

for all $v^S \in T^S$, with strict inequality for at least one $v^S$. If, for each pair $v^B$, $\hat{v}^B$, there is at least one value $\bar{v}^S$ such that

$$(v^B - \hat{v}^B)[y(v^B,\bar{v}^S) - y(\hat{v}^B,\bar{v}^S)] \geq 0$$



then truth can not be dominated and y can be rationalized for most x. Myerson-Satterthwaite's structure plus incentive compatibility requires this inequality to hold for all $\bar{v}^S$. Our result shows that it need hold only for one $\bar{v}^S$, which can differ for different pairs $v^B$, $\hat{v}^B$. This fact generalizes to multi-unit auctions (Gresik-Satterthwaite [4]) and provides the logical foundation, along with the revelation principle, for the observation that behavior in oral double auctions which satisfies the Easley-Ledyard [1] hypotheses is rationalizable as Bayesian equilibrium behavior consistent with preferences which are increasing in money.

The above results for c-mechanisms can be easily generalized to cover all mechanisms with the following definition.

Definition: Given $\Pi : T \to \Delta(D)$ and preferences w, we say that truth is dominated in $\Pi$ for w if there is an i and $t^i$, $s \in T^i$, $t^i \neq s$ such that

$$\sum_{d \geq d'} \mu(d, t_{-i}, t_i) \leq \sum_{d \geq d'} \mu(d, t_{-i}, s)$$

for all $d', t_{-i}$, with strict inequality for at least one $(d', t_{-i})$.

That is, $\Pi(t/s)$ provides a probability measure on D which stochastically dominates (using $w^i$) the probability measure if truth, $t^i$, is used. This definition is entirely consistent with that given earlier and all results generalize in the obvious manner.

In conclusion, any direct revelation mechanism for which truth is not a dominated strategy can be rationalized. This means that,

given any preferences and any mechanism, any behavior which is not dominated for those preferences can be rationalized in a way consistent with those preferences, and any behavior which is not weakly dominated can be strictly rationalized.

IV.4. Restrictions on Preferences and Beliefs

One should be interested whether the weak conditions, developed in the previous sections, for rationalizing performance, rely mainly on the ability to manipulate utility or on the ability to manipulate beliefs. It is easy to show that it is indeed the former. If we are allowed to choose any $u^i$ and, in particular, a $u^i$ which may depend on $t_{-i}$, then even if we are constrained to a single vector of beliefs, we can rationalize almost anything. This follows from an observation in Myerson [17].

Theorem 6 If (a) $\Pi$ is an icdr in $a$, and (b) $(p^{*1},\ldots,p^{*n})$ are priors on $T_{-i}$ for $i = 1,\ldots,n$ such that $p^{*i}(t)) > 0$ whenever $u^i(d,t)p^i(t) > 0$, then $\Pi$ is an icdr in $a^*$ where $a^* = (D,T,u^*,p^*)$ and

$$u^{*i}(d,\ t) = \begin{cases} [u^i(d,t)p^i(t)]/p^{*i}(t) & \text{if } p^{*i}(t) \neq 0, \\ 0 & \text{otherwise} \end{cases}$$

Thus, given $(p^{*1},\ldots,p^{*n})$ with $p^{*i}(t) > 0 \quad \forall i, t$, (we call such $p^*$ positive priors), $\Pi$ is rationalizable, constrained by $w$, iff $\Pi$ is rationalizable constrained by $w$ and $(p^{*1},\ldots,p^{*n})$.

Proof: Straightforward.



Thus you give me a performance function, $\Pi$, preferences, $w$, and positive priors $p^*$. If truth is not dominated in $\Pi$ for $w$ then I can always find utilities, $u^*$, consistent with $w$, which rationalize $\Pi$ as the outcome of Bayesian equilibrium behavior.

Stated another way: for any game with positive priors and any strategies which are not dominated there is a monotonic increasing transformation of the utilities such that those strategies are a Bayes equilibrium in the transformed game.

IV.5. Restrictions on Utility

To now we have seen that even if preferences and beliefs are prespecified, almost any performance function can be incentive compatible in some environment. On the other hand, we know from Gresik-Satterthwaite [4], Milgrom-Weber [14], Myerson-Satterthwaite [18], and Wilson [21] that that the joint assumptions of independent beliefs, transferable utility, risk neutrality and incentive compatibility do restrict possible performance. Therefore, we should not expect to be able to rationalize all performance if utilities are prespecified. Much can, however, still be done.

Our first result provides the basis for the rest of this section.

Theorem 7: Let $\Pi : T \to \Delta(D)$, with D and T finite and let $\mu(d,t)$ be the probability assigned by $\Pi(t)$ to $d \in D$. Let the utilities $w = (w^1,\ldots,w^n)$ be given where $w^i: D \times T \to R^1$ for all i.

There exist priors $p^* = (p^{*1},\ldots,p^{*n})$, where

$p^{*i}(\cdot, t_i) \in \Delta(T^{-i})$, such that $\Pi$ is an icdr mechanism in $\alpha = (D,T,w,p^*)$ if and only if there do not exist $y^i_{\ell s} \geq 0$ for each i, and all $\ell$, $s \in T^i$ such that

$$\sum_s y^i_{\ell s}[\sum_d w^i(d,t/\ell)(\mu(d,t_{-i},\ell) - \mu(d,t_{-i},s))] < 0 \qquad (6)$$

for all $t_{-i} \in T^{-i}$

Proof $\Pi$ is an icdr mechanism in $(w,p^*)$ if and only if for all i, and all $\ell$, $s \in T^i$,

$$\sum_{t_{-i}} \sum_d w^i(d,t_{-i},\ell)[\mu(d,t_{-i},\ell) - \mu(d,t_{-i},s)] \; p^{*i}(t_{-i},\ell) \geq 0.$$

Now apply Gale [2, Lemma 2.10], another variation of Farkas' Lemma, to get the result.

(QED)

As it stands, this is not very informative. Some sufficient conditions may provide a little help.

Corollary 7.1: Given the c-mechanism $\Pi : T \to D$, and the utilities $w = \langle w^1,\ldots,w^n\rangle$. If, for each i and each $t^i \in T^i$, $\exists\ \hat{t}_{-i}$ such that, for all $s \in T^i$,

$$w^i[\Pi(\hat{t}_{-i},t_i),\hat{t}_{-i},t_i] \geq w^i[\Pi(\hat{t}_{-i}/s),\hat{t}_{-i},t_i], \qquad (7)$$

then $\exists\ p^*$ such that $\Pi$ is incentive compatible in $\alpha = (D,T,w,p^*)$.

Proof: Given such a $\hat{t}_{-i}$ it is impossible to satisfy (6) with non-negative $y^i_{t_i s}$.

(QED)



Thus, if there is at least one possible vector of others' types such that i would not want to misrepresent, when faced with those types, we can provide beliefs such that i will never want to misrepresent. One such obvious prior is $p^{*i}(\hat{t}_{-i},t_i) = 1$. If inequality (7) is strict, it is also possible to find an appropriate $p^{*i}$ such that $p^{*i}(t_{-i},t_i) > 0$ for all $t_{-i}$. Thus, generally we could restrict our attention to positive priors.

Remark: This result is reminiscent of the results in Ledyard [10] where it is shown that, when the mechanism can not depend on common knowledge beliefs, incomplete information can not make a mechanism incentive compatible. Here, even if the mechanism can depend on common knowledge beliefs, a similar result obtains in one direction. If $\Pi$ is incentive compatible under complete information in $(\hat{t}_{-i},t_i)$ then $\exists\ p^{*i}$ such that $\Pi$ is incentive compatible in the incomplete information environment since if i believes $\hat{t}_{-i}$ is highly probable then i will always use the strategy $t^i$.

Corollary 7.2: Corollary 7.1 can be applied to p-mechanisms if (7) is replaced by, $\forall s$,

$$\sum_d w^i(d,\hat{t}_{-i},t_i)\mu(d,\hat{t}_{-i},t_i) \geq \sum_d w^i(d,\hat{t}_{-i},t_i)\mu(d,\hat{t}_{-i},s).$$

Proof: Same as Corollary 1.

(QED)

A slightly more general, but still obvious, sufficient condition can be given.

Corollary 7.3: If

$\forall i$ and $t_i \varepsilon T^i$ $\exists$ $B \subseteq T^{-i}$ such that

$$\sum_B \sum_d w^i[d,t][\mu(d,t) - \mu(d,t/s)] \geq 0 \quad \forall s \varepsilon T^i,$$

then $\exists p^*$ such that $\Pi$ is incentive compatible in $\alpha = (D,T,w,p^*)$.

Proof: Let

$$p^{*i}(t) = \begin{cases} \frac{1}{||B||} & \text{if } t_{-i} \varepsilon B \\ 0 & \text{otherwise} \end{cases}$$

(QED)

Using the Myerson-Satterthwaite example with $w^i(dt) = vy(t) - x(t)$, one can see that if beliefs can be type dependent then many outcome rules, $\langle y(\cdot),x(\cdot)\rangle$, can be rationalized. For example, there is no necessity for $\sum_{t-i} y(t_{-i},v)$ to be non-decreasing in the buyer's value, v.

For completeness, let us consider one further restriction to independent beliefs.

Theorem 8: Let $\Pi : T \to \Delta(D)$ with D and T finite and let $\mu(d,t)$ be the probability assigned d by $\Pi(t)$. Let the utilities $w = (w^1,\ldots,w^n)$ be given.



There exist priors for each i, $p^{*i} \varepsilon \Delta(T^{-i})$, independent of $t_i$, such that $\Pi$ is incentive compatible in $\alpha = (D,T,w,p^*)$ if and only if $\nexists$ $y^i_{\ell s} \geq 0$ $\forall i$ and every $\ell,s \varepsilon T^i$ such that, for all $t_{-i} \varepsilon T^{-i}$,

$$\sum_s \sum_\ell \sum_d w^i(d,t_{-i},\ell)[\mu(d,t_{-i},\ell) - \mu(d,t_{-i},s)] \quad y^i_{\ell s} < 0.$$

Proof: Virtually identical to the previous proof.

(QED)

Sufficient conditions can also be provided for this case although they are becoming increasingly uninformative. For example if there is a $t_{-i}$ such that

$$\sum_d w^i(d,t_{-i},\ell)[\mu(d,t_{-i},\ell) - \mu(d,t_{-i},s)] \geq 0$$

for all $s \varepsilon T^i$ and all $\ell \varepsilon T^i$ then $\Pi$ can be rationalized. Obviously, this is not close to being a necessary condition.

It is highly likely that a complete characterization of rationalizable $\Pi$ given prior restrictions on utilities is not possible and each specific utility restriction must be handled on a case by case basis. Some success has been achieved by others in this direction. One example is the case of private values and independent beliefs. Another is the case of affiliated values. The reader can surely provide others.[5]

## V. SUMMARY AND OPEN QUESTIONS

There are at least three conclusions which one should reach as a result of the above theorems:

(1) Any observed behavior can be rationalized as the outcome of a Bayesian equilibrium of some game.

(2) Even if types, outcomes, preferences on outcomes, positive prior beliefs, and an informationally consistent game form are all specified a priori, all strategic behavior which is not dominated can be rationalized as the outcome of a Bayesian equilibrium of a game with the given game form, those priors and utilities consistent with those preferences.

(3) Only when one places severe a priori restrictions on the functional forms of utilities and priors does one get any meaningful restrictions on behavior. The assumptions of risk-neutral quasi-linear preferences and independent values are very special.

These conclusions are similar to those, from Walrasian general equilibrium theory, which apply to aggregate excess demand functions. (See Sonnenschein [20]). They do not imply that the hypothesis of Bayesian equilibrium has no content; rather that it is only the combination of assumptions on utilities, priors, and equilibrium which jointly provide meaningful implications.

Several other observations can also be made. First, a full characterization of the triples of utilities, priors, and performance, such that that performance is incentive compatible for those environments, remains to be done. Second, among the class of incentive compatible direct mechanisms are those which are efficient; that is, either those one might expect to arise as the result of negotiation prior to play, or those which are durable. Although it



was unnecessary to consider efficiency when explaining observed behavior, it would be nice to know whether the hypothesis of efficiency places any additional meaningful constraints on behavior. That is, "what behavior can be rationalized as both incentive compatible and efficient?". My conjecture is that the conclusions will be similar to those above. Although a full analysis remains for future research, a simple observation may explain the basis for the conjecture. As is explained in Myerson [17] and Wilson [22] there are several concepts of efficiency. The two most commonly used are ex-ante and interim. The key observation is that, for this line of research, the two are equivalent. That is, a particular performance function will be an interim efficient direct revelation mechanism in an environment if and only if there is a linear transformation of utilities such that that performance function is also an ex ante efficient mechanism in the transformed environment.[6] Since the transformation is linear it is, in the language of Myerson [17], evaluation-equivalent. Thus Bayesian equilibrium and incentive compatibility are unchanged. This means that behavior can be rationalized as interim efficient if and only if it can be rationalized as ex ante efficient. (For fixed utility, ex ante implies interim but not conversely. For fixed behavior, they are equivalent.) Although it remains to be established, I suspect that only a fairly simple extra condition is needed to show that if there is an environment such that a given performance function is an incentive compatible direct revelation mechanism then there is a

reasonable transformation of the utilities in that environment such that that behavior is interim efficient, and therefore, ex-ante efficient.

As a third observation, none of the results in this paper should be taken as diminishing the normative content of Bayesian behavior in game-theoretic situations. Bayesian equilibrium remains a powerful way of thinking about strategic situations in which players have incomplete information. It is not unreasonable that players should play that way. However, in spite of the validity of the normative content of Bayesian equilibrium, I think one now has to be very careful of the conclusions one draws from models which employ Bayesian behavior as a positive theory. Since virtually any behavior can be rationalized with some utility function, any behavior can be justified as the result of a Bayesian equilibrium. Implications of models using specific utilities and priors need not necessarily be applicable to all data. For example, combined with the revelation principle, the work of Gresik and Satterthwaite [4] tells us[7] that, in auctions involving multiple units, if there is risk neutrality, transferable utility, and independent uniform priors then no extra marginal units will be traded in say, oral double auctions under Bayesian behavior. The experimental evidence is strongly to the contrary. (See Easley and Ledyard [1]). Does this mean that the observed behavior is unreasonable or that the hypothesis of Bayesian equilibrium is inapplicable? Obviously not. It strongly suggests either that risk attitudes are important or that beliefs may not be



independent. The specific functional assumptions are probably at fault, not the theoretical structure. We must look elsewhere for principles to explain these observed behaviors.[8] It is an open question whether there are natural restrictions on utility which prevent the uninformative explanation that all experimental and market generated observations are Bayesian equilibrium outcomes.

The Bayesian approach to analyzing strategic situations with incomplete information remains a compelling model. Nevertheless, care must be exercised so that conclusions are drawn from general principles and not just from examples.[9] It is highly likely that a close fit to the data and transferability across institutions may ultimately be the only real test of the theory.

FOOTNOTES

* This paper is an extensive revision of one presented in the Theory workshop at Caltech and at the Institute for Mathematics and Its Applications at the University of Minnesota. I thank both sets of participants for their suggestions and comments. I especially thank Kim Border and Charles Plott who asked the right questions at the right time. The revision of the first version has benefited from the insightful and careful comments of Steve Matthews, Mark Satterthwaite, Robert Wilson and a good referee. Finally, I thank the intransigent referee of Easley-Ledyard [1], whose misguided refusal to consider the behavior proposed in that as reasonable led me to the research reported here.

1. When there is more than one Bayesian equilibrium, we will consider a suitable generalization of $\Pi^*$.

2. Another set of papers by Matthews [11], Milgrom [13], Milgrom and Weber [14] and Wilson [21] pick a particular $\langle M,g \rangle$ and ask whether its performance is good. For example, some ask whether $\Pi(\ ;n)$ is an icdr as $n \to \infty$. See Wilson [22] for other references.

3. M. K. Richter has coined the phrase "revealed game theory" to describe this type of issue.



4. B.1 is needed to insure that B.2 is not vacuous.

5. The referee has pointed out that "investigation of weak restrictions on utility, such as risk aversion, might . . . be quite interesting." I agree, but have no specific thoughts or conjectures.

6. Simply use the welfare weights associated with the interim efficiency.

7. The main point of their work is the convergence of efficient auctions to competitive equilibrium as the number of players grows. I am refering to only a subset of that work.

8. For example, the apparently ad hoc model in Easley and Ledyard [1] closely corresponds to observed behavior but does not seem sensible to some economists. The results in this paper, however, suggest that behavior consistent with the Easley-Ledyard assumptions can be always be rationalized as the Bayesian Equilibrium of an Oral Double Auction.

9. To paraphrase Robert Wilson's closing remarks in Wilson [23], Bayesian behavior can explain lots of behavior; maybe too much.

REFERENCES


1. David Easley and John Ledyard, "A Theory of Price Formation and Exchange in Oral Auctions," Center for Mathematical Studies in Economics and Management Science, Discussion Paper No. 461, Northwestern University, (rev.) July 1982.

2. David Gale, "The Theory of Linear Economic Models," McGraw Hill, New York, 1960.

3. A. Gibbard, Manipulation of voting schemes: A general result," Econometrica 41 (July 1973), 587-602.

4. Thomas A. Gresik and Mark A. Satterthwaite, "The Number of Traders Required to Make a Market Competitive: The Beginnings of a Theory," Center for Mathematical Studies in Economics and Management Science, Discussion Paper No. 551, Northwestern University, February 1983.

5. Milton Harris and Artur Raviv, Allocation mechanisms and the design of auctions, Econometrica 49 (November 1981), 1477-1499.

6. Milton Harris and Robert M. Townsend, Resource allocation under asymmetric information, Econometrica 49 (January 1981), 33-64.

7. L. Hurwicz, Optimality and informational efficiency in resource allocation processes, in "Mathematical Methods in the Social Sciences" (K. J. Arrow, S. Karlin, and P. Suppes, Eds.), Chapt. 3, Stanford University Press, Stanford, 1960.





8. L. Hurwicz, On informationally decentralized systems, in "Decision and Organization" (C. B. McGuire and R. Radner, Eds.), North Holland Press, Amsterdam, 1972.

9. J. Laffont and E. Maskin, The theory of incentives: An overview, in "Advances in Economic Theory" (Werner Hildebrand, Ed.), Chapt. 2, Cambridge University Press, Cambridge, 1982.

10. J. Ledyard, Incentive compatibility and incomplete information, in "Aggregation and Revelation of Preferences" (J. J. Laffont, Ed.), North Holland Press, Amsterdam, 1979.

11. Steven Matthews, "Risk Aversion and the Efficiency of First- and Second-Price Auctions," CCBA Working Paper No. 586, University of Illinois, 1979.

12. Steven Matthews, Selling to risk averse buyers with unobservable tastes, J. Econ. Theory 30 (1983), 370-400.

13. Paul R. Milgrom, A convergence theorem for competitive bidding with differential information, Econometrica 47 (1979), 679-688.

14. Paul R. Milgrom and Robert J. Weber, A theory of auctions and competitive bidding, Econometrica 50 (September 1982), 1089-1122.

15. K. Mount and S. Reiter, The informational size of message spaces, J. Econ. Theory 8 (1974), 161-191.

16. Roger B. Myerson, Optimal auction design, Mathematics of

Operations Research 6 (February 1981), 58-73.

17. Roger B. Myerson, "Bayesian equilibrium and incentive compatibility: An introduction," Center for Mathematical Studies in Economics and Management Science, Discussion Paper No. 548, Northwestern University, February 1983.

18. Roger B. Myerson and Mark Satterthwaite, "Efficient mechanisms for bilateral trading," Center for Mathematical Studies in Economics and Management Science, Discussion Paper No. 469S, Northwestern University, 1981; forthcoming in J. Econ. Theory, May 1984.

19. A. Postlewaite and D. Schmeidler, Revelation and implementation under differential information, in "Essays in Honor of Leonid Hurwicz", University of Minnesota Press, Minneapolis, forthcoming.

20. H. Sonnenschein, Do walras' identity and continuity characterize the class of community excess demand functions? J. Econ. Theory 6 (1973), 345-354.

21. Robert B. Wilson, "Double auctions," IMSSS Technical Report 391, Stanford University, 1982. To appear in "Essays in Honor of Leonid Hurwicz," University of Minnesota Press, Minneapolis.

22. Robert B. Wilson, "Incentive efficiency of double auctions," IMSSS Technical Report 431, Stanford University, October 1983.



23. Wilson, Robert B. "Reputations in Games and Markets," IMSS Technical Report 434, Stanford University, November 1983.